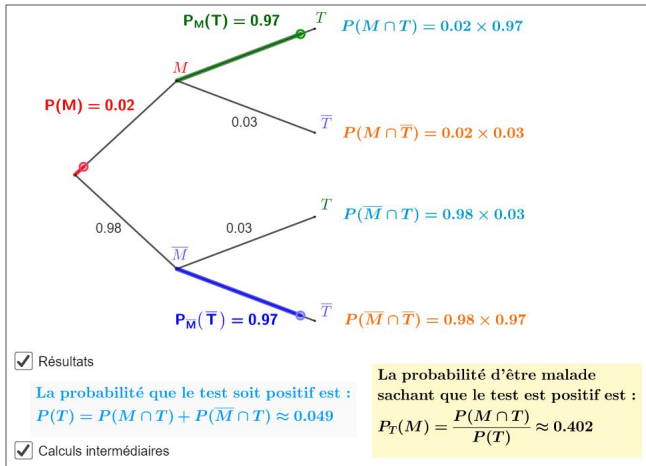


- L'appliquette GeoGebra « inférence bayésienne et dépistage avec un arbre » permet d'illustrer l'influence de la sensibilité, de la spécificité et de la prévalence sur les VPP et VPN.

media.eduscol.education.fr/other/les_mathematiques_de_l_ES/15/1/RA19_T_ESmaths_3.5.inference_bayesienne_depistage_arbre_1227151.ggb



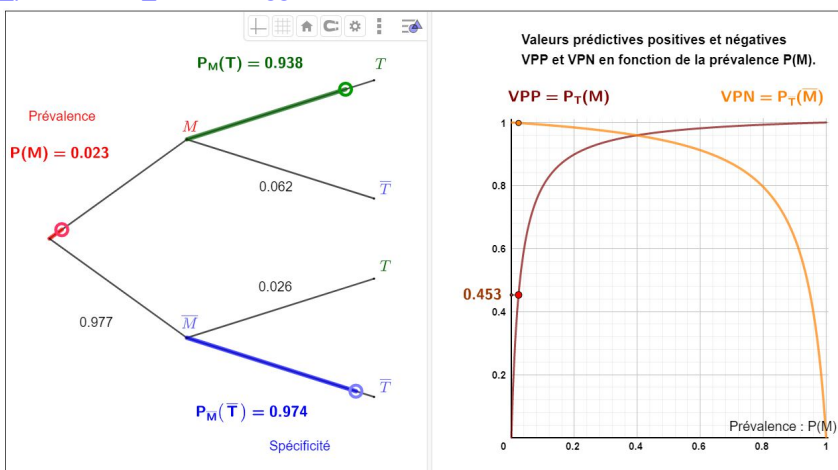
- Tableau récapitulatif – cas général

S_e est la sensibilité, S_p est la spécificité

	Test positif (T^+)	Test négatif (T^-)	Total
Malades (M)	$P(T^+ \cap M) = P(M) \times P_M(T^+)$ $= p \times S_e$	$P(T^- \cap M) = P(M) \times P_M(T^-)$ $= p \times (1 - S_e)$	$P(M) = p$
Non malades (\bar{M})	$P(T^+ \cap \bar{M}) = P(\bar{M}) \times P_{\bar{M}}(T^+)$ $= (1 - p) \times (1 - S_p)$	$P(T^- \cap \bar{M}) = P(\bar{M}) \times P_{\bar{M}}(T^-)$ $= (1 - p) \times S_p$	$P(\bar{M}) = 1 - p$
Total	$P(T^+) = p \times S_e + (1 - p) \times (1 - S_p)$	$P(T^-) = p \times (1 - S_e) + (1 - p) \times S_p$	1

- L'appliquette GeoGebra « Valeur prédictive en fonction de la prévalence » permet de visualiser différentes courbes de valeurs prédictives en fonction des caractéristiques du test.

media.eduscol.education.fr/other/les_mathematiques_de_l_ES/16/1/RA19_T_ESmaths_3.5.valeur_predictive_fonction_prevalence_1227161.ggb



● Lien OMS pour en savoir plus sur le virus Ebola :
<https://www.who.int/fr/news-room/fact-sheets/detail/ebola-virus-disease>

● [Tangente Hors série n°58, pp. 8 à 10]

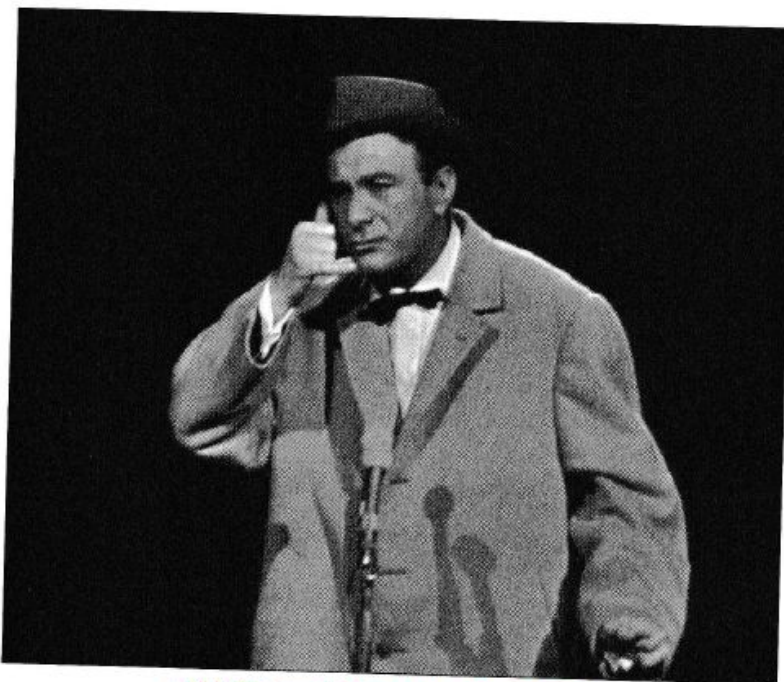
Diagnostic

Les raisonnements médicaux

Les raisonnements mathématiques débouchent sur des certitudes, pour peu qu'ils respectent la logique et l'axiomatique, tandis que le plus compétent des médecins peut commettre une erreur dans son diagnostic.

Comment expliquer cette différence ?

Un mathématicien, c'est bien connu, ne se trompe jamais... pour autant qu'il reste constamment dans la théorie étudiée et en respecte les règles. Schématiquement, tout raisonnement mathématique peut se ramener à un simple syllogisme faisant intervenir deux propositions, à savoir l'hypothèse H , ensemble des informations que l'on suppose connues, et la thèse T que l'on cherche à démontrer : si l'on suppose vraies l'hypothèse H et l'implication de la thèse connaissant l'hypothèse (qui est notée $H \Rightarrow T$), alors la thèse T est également vraie.



« Dis tonton, pourquoi tu tousses ? », par Fernand Raynaud.

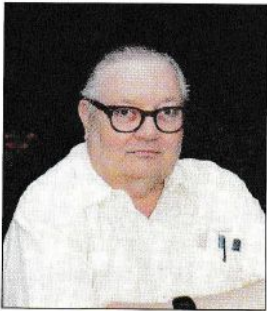
✦ Du « certain » au « plausible »

De manière schématique (voire caricaturale), on peut affirmer qu'un médecin, lorsqu'il pose un diagnostic, raisonne aussi en trois phases :

1. Il observe chez son patient un symptôme S (par exemple, le malade tousse).
2. D'après ses connaissances médicales, il sait qu'une maladie M entraîne souvent le symptôme S observé (en poursuivant le même exemple, le docteur sait qu'une infection pulmonaire entraîne généralement la toux).
3. Il en conclut, de manière plausible, que le patient examiné a contracté la maladie M (sur le même exemple, le patient souffre vraisemblablement d'une infection pulmonaire).

Si le raisonnement est en apparence semblable à celui du mathématicien (à condition de remplacer H par S et T par M), ce n'est en réalité pas le cas : il en diffère par le fait que l'implication considérée ici est « inversée », ce n'est pas $S \Rightarrow M$, mais $M \Rightarrow S$, équivalente à $\text{Non } S \Rightarrow \text{Non } M$. Techniquement, le médecin utilise la réciproque de l'implication du mathématicien. Cette différence est fondamentale, car elle a pour effet de rendre la conclusion non plus certaine, mais seulement plausible.

Le physicien américain Richard Threlkeld Cox (1898–1991) a « quantifié » l'idée de *plausibilité*. Il a même établi un théorème affirmant, intuitivement, que si les plausibilités satisfont un ensemble (restreint) d'hypothèses (somme toute naturelles), alors leur emploi se ramène en fait à l'utilisation du calcul des probabilités. Cet énoncé est désormais connu sous le nom de *théorème de Cox-Jaynes*, le second auteur étant le mathématicien américain Edwin Thompson Jaynes (1922–1998) qui a largement contribué à la diffusion de ce résultat important. Notons également que cet énoncé est exploité dans les travaux de recherche de Stanislas Dehaene (né en 1965), mathématicien, psychologue et neu-



De haut en bas, Richard Threlkeld Cox, Edwin Thompson Jaynes et Stanislas Dehaene.

roscientifique réputé (voir *Tangente* 159). Pour plus d'informations, on pourra consulter son site à l'adresse www.college-dc-france.fr/site/stanislas-dehaene.

À la lumière de ce qui précède, réexaminons le « syllogisme médical » énoncé ci-dessus.

La conclusion de ce raisonnement est la suivante : étant donné la présence du symptôme S , il est probable qu'il s'agisse de la maladie M . En d'autres termes, la plausibilité de la conclusion est mesurée par la probabilité de M sachant S , c'est-à-dire par la probabilité conditionnelle $P(M|S)$.

Or, la théorie des probabilités livre cette égalité (voir l'encadré) :

$$P(M|S) = [P(S|M)P(M)]/P(S).$$

Puisque le dénominateur ne fait pas intervenir la maladie M , on en conclut que le verdict recherché peut aussi se traduire par la relation de proportionnalité (le symbole \sim signifiant ici « est proportionnel à »)

$$P(M|S) \sim P(S|M)P(M).$$

Dans cette formule interviennent trois nombres :

- $P(M|S)$ est la probabilité *a posteriori* de M , en ce

Quelques rappels sur les probabilités conditionnelles

Sur l'ensemble des événements d'un univers donné, considérons une probabilité P et deux événements A et B de probabilité non nulle. Les probabilités, qualifiées de *conditionnelles*, de A sachant B et de B sachant A sont respectivement définies par :

$$P(A|B) = P(A \cap B)/P(B) \text{ et } P(B|A) = P(B \cap A)/P(A).$$

Comme $A \cap B = B \cap A$,

la double égalité suivante résulte de ces définitions :

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A).$$

De là, on déduit aisément que $P(A|B) = [P(B|A)P(A)]/P(B)$.

Pour approfondir cette question, voir notamment l'article sur le théorème de Bayes dans Tangente Sup 73-74.

sens qu'elle est une déduction logique de la maladie après l'observation de S ; elle est parfois appelée dans le monde médical la *valeur prédictive positive* du symptôme pour la maladie ;

- $P(S|M)$ est la *sensibilité diagnostique* du symptôme (pour la maladie M) ;
- $P(M)$ est la *probabilité a priori* de M (qui peut notamment dépendre de l'environnement de travail du médecin) ou plus techniquement la *prévalence* de la maladie M .

Le premier de ces nombres (celui que le médecin souhaite communiquer au malade) est donc proportionnel au produit des deux autres. Or, ces derniers sont plus faciles à déterminer que la probabilité recherchée ; ils font en effet partie des connaissances médicales ou peuvent être estimés assez aisément. Pour l'exemple de départ, le médecin sait qu'une personne atteinte d'infection pulmonaire est souvent amenée à tousser et qu'il y a, au moment de l'examen, une proportion importante d'infections pulmonaires dans la population concernée, c'est-à-dire que la sensibilité diagnostique et la prévalence sont élevées (donc, proches de 1). Dès lors, son verdict est donné sans hésitation, en admettant bien sûr que le docteur observe que le symptôme (ici la toux) possède une probabilité qui, par souci de simplicité, est ici assimilée à l'unité : la probabilité *a posteriori* est grande et il s'agit presque certainement d'une infection pulmonaire !

+ Quand il y a hésitation entre plusieurs diagnostics

La situation est en réalité plus complexe car le médecin est en général amené à opérer un choix parmi deux ou plusieurs maladies pour un même symptôme S .

Supposons qu'il envisage une autre maladie M^* pour le même symptôme S . Si, comme illustration, on reprend le cas de la toux (symptôme S), une autre origine peut aussi être le cancer du poumon (maladie M^*) : dans ce cas, la sensibilité diagnostique $P(S|M^*)$ est également élevée, mais la prévalence $P(M^*)$ est basse dans l'esprit du médecin (car, par exemple, le sujet examiné est jeune et ne fume pas) ; de la sorte, le produit du nombre $P(S|M^*)$, qui est proche de 1, par le nombre $P(M^*)$, qui, lui, est presque nul, est faible.

Au surplus, rappelons que le produit correspondant relatif à la maladie M est proche de 1. En conséquence, si l'on considère le rapport des deux probabilités *a posteriori* (car on veut comparer la plausibilité des deux maladies), on trouve

$$\frac{P(M|S)}{P(M^*|S)} = \frac{[P(S|M)P(M)]/P(S)}{[P(S|M^*)P(M^*)]/P(S)} = \frac{P(S|M)P(M)}{P(S|M^*)P(M^*)}.$$

La simplification opérée rend superflu le dénominateur commun du terme central, à savoir $P(S)$, lors d'une telle comparaison. En conséquence, dans le cas présent, la maladie M est plus plausible que M^* (ou, une infection pulmonaire est nettement plus probable qu'un cancer du poumon) : en effet, le rapport des deux probabilités *a posteriori* cherchées est visiblement supérieur à 1.

Comme le montre l'exemple précédent, dans les raisonnements médicaux, le praticien suspecte généralement la maladie pour laquelle la probabilité *a posteriori* est la plus grande, et ceci parmi toutes les maladies en relation potentielle avec le symptôme observé : c'est ce qu'on appelle dans le jargon scientifique anglo-saxon, le *MAP* (« *maximum a posteriori hypothesis* »).



+ Les trois médecins

Mais ce critère peut être délicat à mettre en œuvre dans la pratique. Sans approfondir la question, on peut illustrer ce point par l'exemple ci-dessous, construit fictivement (par souci didactique) avec quelques données simples.

Trois médecins, numérotés de 1 à 3, sont confrontés à un même symptôme observé S .

Ils hésitent, assez spontanément, entre deux maladies possibles M ou M^* , mais avec des points de vue un peu différents. Voici leurs calculs :

Médecin 1 :

$$P(S|M) = 0,8 ; P(M) = 0,8 ;$$

$$P(S|M^*) = 0,9 ; P(M^*) = 0,7.$$

Pour ce premier médecin, la sensibilité diagnostique du symptôme S est plus grande pour la maladie M^* que pour M (puisque $0,9 > 0,8$) et pourtant, en vertu du critère *MAP*, la maladie M va être diagnostiquée ; en effet,

$$P(S|M)P(M) = 0,8 \times 0,8 = 0,64 \text{ alors que}$$

$$P(S|M^*)P(M^*) = 0,9 \times 0,7 = 0,63.$$

Médecin 2 :

$$P(S|M) = 0,9 ; P(M) = 0,7 ;$$

$$P(S|M^*) = 0,8 ; P(M^*) = 0,8.$$

Pour ce deuxième médecin, la situation est inversée par rapport à celle de son premier confrère : la sensibilité diagnostique du symptôme S est plus élevée pour la maladie M , mais finalement c'est la maladie M^* qui va être retenue (toujours en vertu de *MAP*).

Médecin 3 :

$$P(S|M) = 0,8 ; P(M) = 0,9 ;$$

$$P(S|M^*) = 0,9 ; P(M^*) = 0,8.$$

Ce troisième docteur, que va-t-il diagnostiquer puisque les probabilités *a posteriori* de M et de M^* coïncident ?

À la lumière de ce petit texte introductif, nos lecteurs peuvent entrevoir la difficulté qu'ont les médecins de réaliser un diagnostic aussi fiable et incontestable que l'est un raisonnement mathématique.

□— A. A. & J. B.